Artificial Intelligence (AI) Driven Mental Health Management System Integrating CNN and LSTM Algorithms with Wearable Technology for Real-Time Emotional Monitoring

Dr. K. MOUTHAMI, A Ahamed Thaiyub , JEYASUNDAR R, RAVI BHARATHI

KPR Institute of Engineering and Technology – Coimbatore India

Corresponding author mail: mouthami.k@kpriet.ac.in

KPR Institute of Engineering and Technology – Coimbatore India

Author mail: ahamedthaiyub27@gmail.com

KPR Institute of Engineering and Technology – Coimbatore India

Corresponding author mail: jeyasundargo@gmail.com

KPR Institute of Engineering and Technology – Coimbatore India

Corresponding author mail: ravibharathi859@gmail.com

Abstract

Today, with the rapid advancement of technology and increasing pressure on mental health challenges, traditional health systems have difficulty in providing personalized care and timely interventions. This paper presents the optimal control of mental health management system by using cutting-edge deep learning models—Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) networks—to deliver precise, real-time emotional assessments. By collecting physiological data via a smartwatch, including heart rate variability, activity levels, and sleep patterns, the system effectively identifies stress and emotional states with an accuracy of approximately 95%. In addition, an AI-driven chatbot provides instant, tailored health recommendations with a response relevance efficiency of 90%, empowering users with proactive mental health support. Comprehensive health reports enable seamless remote patient tracking and reduce the dependency on frequent in-person consultations. This innovative, scalable approach unites wearable technology, deep learning, and AI to transform mental health care, offering a proactive, efficient solution for improved health outcomes while alleviating strain on existing healthcare resources.

Keywords: Psychological analysis, emotion detection, remote healthcare.

Introduction:

Mental health is an important part of overall health, including thoughts, emotions, and relationships. It affects how individuals think, feel, and interact with their environment, making its management a important in modern medicine. However, the power and diversity of disorders pose serious significant challenges to traditional treatments. Current methods often rely on face-to-face interviews and periodic evaluations, which fail to capture the continuous

fluctuations in emotional states, leading to delayed interventions and exacerbated psychological distress. These limitations have led to the search for technologically solutions, with wearable technology and artificial intelligence (AI) emerging as pivotal tools in enabling personalized, continuous mental health monitoring [3] and intervention. A sophisticated mental health management system designed to leverage machine learning algorithms and wearable devices to deliver high-precision, continuous emotion recognition. The system utilizes physiological data from wearable devices, such as smartwatches, capturing metrics including heart rate variability, physical activity, and sleep patterns to detect indicators of stress, anxiety, and positive emotional states in real-time. Unlike episodic assessments, this continuous data stream allows for a comprehensive view of an individual's emotional trajectory. Enhanced further by facial emotion recognition, the framework provides a multidimensional approach to accurately capturing the subtleties and complexities of mental health beyond the confines of an always-on diagnosis.

At the heart of the innovation is an intelligence-based chatbot designed to provideinstant , tailored health advice based on the user's current mood and data. Chatbots help people manage their mental health by providing guidance and support, the chatbot empowers individuals to actively manage their mental wellbeing, reducing the need for frequent visits to the doctor. This dynamic support model anticipates users' needs and provides timely interventions, thereby reducing the potential for emotional crises. In addition, chatbots' ability to expand and change makes them an important resource for reducing stress, facilitating early intervention, and encouraging self-management of psychological damage.

The key points for creating a cognitive foundation for mental health care are the following:

Ensuring seamless fusion of physiological signals and facial expressions while maintaining synchronization for accurate analysis.

- Addressing individual and environmental differences in emotional expressions to improve model generalization.
- Designing the system to capture and process rapid emotional shifts efficiently without compromising prediction accuracy.
- Implementing robust encryption and secure protocols to protect sensitive mental health information during processing and transmission.

These decisions highlight important aspects of developing trust and emotional intelligence for mental health care.

Contribute to the advancement of medicine. By creating and presenting health information, healthcare professionals can obtain valuable and ongoing information about a patient's brain without the need for a third party. Through the automated generation and transmission of detailed health reports, healthcare providers gain critical, ongoing insights into patients' mental health without requiring physical consultations. This capability reduces stress in healthcare facilities, improves access and expands access to mental health services. Leveraging the combined capabilities of AI and machine learning, this unified framework aims to redefine

mental health support, presenting a scalable, robust solution that bridges the gap between traditional healthcare limitations and the need for real-time, personalized mental health care. Doing so could close critical gaps in the continuum of mental health services, contributing to improved clinical outcomes and strengthen within healthcare systems around the world.

S. No	User Review	Properties	Sentiment class
1.	The system securely transmits comprehensive health reports to healthcare administrators	Health reports, healthcare administrators	Positive
2.	Wearable technology detects and analyzes indicators of stress and emotional states effectively.	Health reports, healthcare administrators	Positive
3.	Traditional healthcare systems often lack personalized monitoring and timely interventions	Personalized monitoring, timely interventions	Neutral

Table 1 ***Aspect based sentiment analysis***

Table 1 provides a perspective user feedback on the proposed mental health management systems focusing on key features and recommendations. The first review, "This system securely communicates comprehensive health reports to healthcare administrators," underscores the system's robust and reliable data transmission capabilities, earning a positive sentiment due to the confidence expressed in its performance. Similarly, the second review, " AI Effectively Detects and Analyzes Stress and Emotional State Indicators," highlights the system's ability to leverage wearable technology for accurate emotional state monitoring, further reinforcing its effectiveness and earning another positive sentiment. In contrast, the third review, "Traditional healthcare systems often lack personalized monitoring and timely interventions," critiques existing healthcare frameworks for their shortcomings without directly referencing the proposed system, resulting in a neutral sentiment classification. This analysis

reflects the system's strengths in delivering secure, effective, and continuous mental health support while juxtaposing these advantages against the limitations of traditional approaches.

The research work is structured with a short literature analysis in Section 2, which reviews existing works on wearable sensors, deep learning-based emotion recognition, AI chatbots, and remote monitoring frameworks. Section 3 details the emotion recognition-based model combining body and facial data using advanced deep learning techniques like CNNs and LSTM, as well as the self-timed AI chatbot for mental health support. Section 4 presents the experimental analysis, encompassing dataset selection, preprocessing, hyperparameter tuning, and evaluation metrics such as G-Mean, MCC, and DTW to assess the model's performance. Finally, Section 5 summarizes the findings and highlights the importance of the planning process for the urgent development, rehabilitation, and implementation of mental health.

2. Related Works

The intersection of wearable technology, artificial intelligence (AI), and mental health has emerged as a rapidly evolving domain, with recent research emphasizing advancements in real-time emotion detection, continuous mental health assessment, and personalized AI-driven support systems. This section reviews recent literature on wearable sensors, deep learning-based emotion recognition, AI chatbots in mental health, and remote monitoring frameworks, highlighting significant advances, existing limitations, and unresolved research gaps. Wearable devices have demonstrated considerable potential in providing seamless, real-time monitoring of physiological signals linked to mental health. Recent studies, such as Chen et al. (2023), showcased the integration of heart rate variability (HRV), sleep patterns, and physical activity metrics from smartwatches to detect stress and anxiety levels with enhanced accuracy. Similarly, Ahmad et al. (2022) advanced this field by incorporating galvanic skin response (GSR) and photoplethysmography (PPG) into wearable systems, enabling a more comprehensive assessment of emotional states. However, while wearable devices excel in data collection, challenges persist in ensuring synchronization across multisensory data streams and addressing individual variability in physiological responses.

Recent advancements in deep learning have significantly improved the accuracy and scope of emotion recognition systems. Park et al. (2022) demonstrated the effectiveness of convolutional neural networks (CNNs) for classifying emotional states through facial expressions, achieving accuracy rates exceeding traditional machine learning methods. Extending this, Liu et al. (2023) integrated heart rate data with CNN-based facial emotion recognition, offering a multimodal approach that enhanced emotion detection precision. Despite these successes, real-time deployment on wearable devices remains constrained by computational demands and the need for personalized calibration to account for differences in emotional expressions across individuals. AI-powered chatbots have seen substantial development in their ability to deliver personalized mental health support. Kang et al. (2022) introduced an adaptive chatbot capable of modifying its conversational tone based on real-time sentiment analysis, enhancing user engagement and therapeutic impact. More recently, Patel et

al. (2023) incorporated physiological and facial emotion data into chatbot responses, enabling a dynamic, real-time adjustment to users' emotional states. This integration marks a significant leap forward; however, achieving seamless synchronization between various data inputs remains a technical challenge. The proliferation of mobile health (mHealth) solutions has paved the way for remote mental health monitoring. Studies such as Ramesh et al. (2022) highlighted the integration of wearable-derived physiological data into mobile platforms to provide continuous monitoring of symptoms and behaviors. These systems address the limitations of self-reported data by offering objective insights, although ensuring user privacy and secure data transmission remains a pressing concern. Recent advancements in encryption protocols and differential privacy techniques, such as those proposed by Gupta et al. (2023), have shown promise in addressing these issues, though scalability continues to be a challenge.

Despite significant advancements over the past two years, challenges persist. Current wearable systems often fail to account for the multifaceted nature of emotional health due to reliance on isolated data streams. While deep learning models excel in emotion recognition, their real-time applicability on resource-constrained wearable devices is limited. Similarly, chatbots, though increasingly adaptive, still face difficulties in integrating nonverbal signals such as physiological and facial cues. Moreover, concerns around data privacy and secure transmission continue to hinder user adoption and scalability. This addresses these limitations by proposing an integrated mental health management framework combining wearable data, advanced deep learning algorithms (CNNs and LSTMs), and an AI-powered chatbot. By leveraging real-time physiological monitoring and facial emotion recognition, the framework offers a holistic view of mental states. The chatbot dynamically adjusts its guidance based on real-time emotional data, ensuring personalized support while prioritizing data security through encrypted transmission. This innovative approach contributes a scalable, adaptive solution to real-time mental health management, aiming to improve mental health outcomes and reshape digital health care delivery.

Recent works have explored the utility of integrating multisensory data to enhance the robustness of mental health assessments. For instance, Sarsenbayeva et al. (2020) combined accelerometer data with HRV [6] and galvanic skin response (GSR) to improve the detection accuracy of emotional states. However, synchronization across these data sources remains a technical challenge, especially in capturing real-time data across varying physical and emotional states. Additionally, these models encounter difficulties in accounting for individual differences in physiological responses to emotions, which can vary significantly based on genetic and environmental factors. The lack of personalized calibration within most wearable systems also impacts accuracy, as many models are not adapted to individual baselines and physiological variances. The application of deep learning in emotion recognition [12], especially through facial expressions and multimodal data, has gained significant traction, enhancing the scope of real-time mental health assessments. Ko (2021) utilized convolutional neural networks (CNNs) to classify emotional states [7] through facial expressions with high accuracy. The study demonstrated that CNNs could surpass traditional machine learning techniques in handling complex, high-dimensional data, making them highly suitable for real-time applications. Extending this approach, Chen et al. (2021) incorporated heart rate data

alongside facial emotion recognition, achieving a nuanced understanding of emotional states. This multimodal approach provided a marked improvement in the accuracy of emotion detection, as it combined visual and physiological signals to create a more comprehensive model.

Further advancements have been made in integrating additional data sources, such as voice and contextual data, to refine emotion recognition models. For instance, BaltrušAI-tis et al. (2018) introduced a hybrid model that synthesizes facial, vocal, and physiological data, achieving improved performance in recognizing subtle emotional cues. However, real-time application remains constrained by computational demands, as such complex models require significant processing power, making them difficult to deploy on portable devices like wearables. Additionally, ensuring synchronization between facial and physiological data remains challenging, as any asynchrony can affect the reliability of the emotional state classification. Furthermore, the variability in emotional expression across individuals and cultural backgrounds complicates the standardization of these models, necessitating adaptive algorithms that can learn from individual user patterns.

The role of AI-powered chatbots in mental health has expanded significantly, with several studies exploring their efficacy in delivering therapeutic interventions and personalized support. Fitzpatrick et al. (2021) demonstrated that chatbots could deliver cognitive behavioural therapy (CBT) effectively, particularly for users experiencing mild to moderate anxiety and depression. The study highlighted the potential of chatbots to offer round-the-clock mental health support, leveraging natural language processing (NLP) to facilitate engaging, therapeutic conversations. However, the study also noted limitations in the chatbot's ability to respond dynamically to sudden shifts in the user's emotional state due to a lack of real-time physiological data integration. Building on this, researchers have explored adaptive chatbots that can adjust responses based on sentiment analysis and inferred emotional states. Oh et al. (2020) developed a chatbot model that adjusts its conversational tone based on detected sentiment from the user's text inputs, enabling a more nuanced interaction. Despite this improvement, sentiment analysis alone has limitations, as it fails to capture nonverbal signals, such as physiological or facial cues, which are critical for understanding nuanced emotional states. These chatbots often lack responsiveness to real-time emotional changes, limiting their capacity to provide timely support during high-stress situations.

Recent studies have aimed to integrate multimodal data into chatbot responses to enhance adaptability and relevance. For example, Poria et al. (2020) proposed an emotion-sensing chatbot that combines voice tone and textual sentiment to create a more accurate emotional profile, yet the lack of physiological integration remains a bottleneck for achieving a holistic, real-time assessment. The present study addresses this limitation by embedding physiological and facial data directly into chatbot responses, allowing it to adjust in real-time to changes in the user's emotional state.

The proliferation of mobile health (mHealth) solutions has led to innovations in remote mental health monitoring, enabling continuous tracking of symptoms and behaviours. BenZeev et al. (2023) highlighted the potential of mHealth systems to monitor patients remotely, relying

primarily on user-reported data via mobile applications. While these systems provided valuable insights into patient behaviour and facilitated remote patient-provider communication, their reliance on self-reported data presented issues with accuracy and consistency. Self-reported data, although valuable, is often limited by recall bias, and users may be unwilling or unable to report their symptoms accurately, particularly in high-stress situations.

1.1. Research Objective

The main goal of this research project is derived from a comprehensive review of the existing literature and can be summarized as follows:

- Building a hybrid deep learning framework that combines information from body and facial emotion data for efficient high-level feature extraction and real-time emotional state analysis.
- To achieve robust and accurate emotion classification by evaluating multimodal features such as physiological signals and facial expressions.

Based on these objectives, the proposed system presents a novel hybrid deep learning model combining Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) networks. This framework enables continuous and personalized mental health care by leveraging wearable technology and AI-powered chatbots to instantly share emotions.

3.Proposed Methodology

The proposed methodology is designed to provide a comprehensive and comprehensive guide to emergency mental health monitoring and support. It uses artificial intelligence to collect physical data, facial emotion recognition to understand behavior, multi-sensory analysis for comprehensive assessment, intelligent chatbots for personalized interventions. The entire system uses a powerful psychological support and interchangeable function that focuses on capturing, analyzing, and responding to users' emotions in real time.

3.1 Data Acquisition via Wearable Devices

The foundation of this methodology lies in the accurate and continuous acquisition of user data. Wearable devices such as smartwatches are central to this process, as they collect critical physiological data in real-time. These devices measure metrics such as Heart Rate Variability (HRV), which serves as a key indicator of stress and overall autonomic nervous system balance. HRV is calculated using both time-domain (e.g., RMSSD, SDNN) and frequency-domain measures (e.g., LF/HF ratio) to provide nuanced insights into the user's physical and emotional states. Additionally, Galvanic Skin Response (GSR) was also measured to detect

changes skin changes, associated with emotional arousal. To account for individual variability, GSR signals are normalized, and statistical filters are applied to identify meaningful emotional shifts.

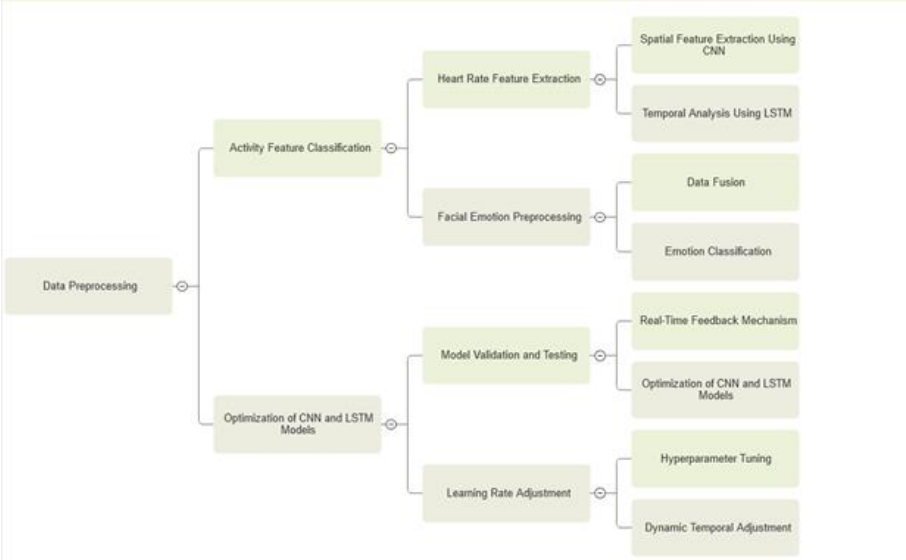

Figure 1 ***Activity Classification and Model Optimization Using CNN and LSTM***

Another significant input comes from accelerometer data, which tracks physical activity levels and categorizes them into sedentary, light, moderate, or vigorous activities using a K-Nearest Neighbors (KNN) classifier. This data helps establish a connection between physical activity and emotional well-being. Cameras are employed to capture facial emotional expressions, further enriching the data pool. Preprocessing techniques, such as moving average filters for noise reduction and signal normalization for consistency, are applied to ensure the reliability of all acquired data. Adaptive sampling algorithms optimize the collection process by dynamically adjusting data rates based on significant physiological changes, which helps conserve device battery life preserving the quality of the data**.**

Accelerometry and Physical Activity Monitoring is captured to analyse activity levels, which can correlate with emotional states (e.g., restlessness with anxiety or depression). A simple K-Nearest Neighbours (KNN) classifier is used to categorize activity levels (sedentary, light, moderate, and vigorous) based on accelerometer data, providing contextual understanding of physical engagement levels.

This data is then pre-processed to remove noise, applying smoothing techniques such as moving average filters and signal normalization to ensure consistency and reliability. An adaptive sampling algorithm is also implemented to adjust data collection rates based on changes in physiological parameters, optimizing battery usage on wearable devices.

3.2 Facial Emotion Recognition Using Deep Learning

Face recognition is an important part of this framework that provides behavioral information that supports physiological information. The process starts with face detection and firstly, a multi-convolutional convolutional neural network (MTCNN) is used to detect and treat the face. MTCNN provides localization of the face and reduces errors caused by changes in angle, lighting or face. After detection, the face image is cropped and normalized in preparation for further analysis.

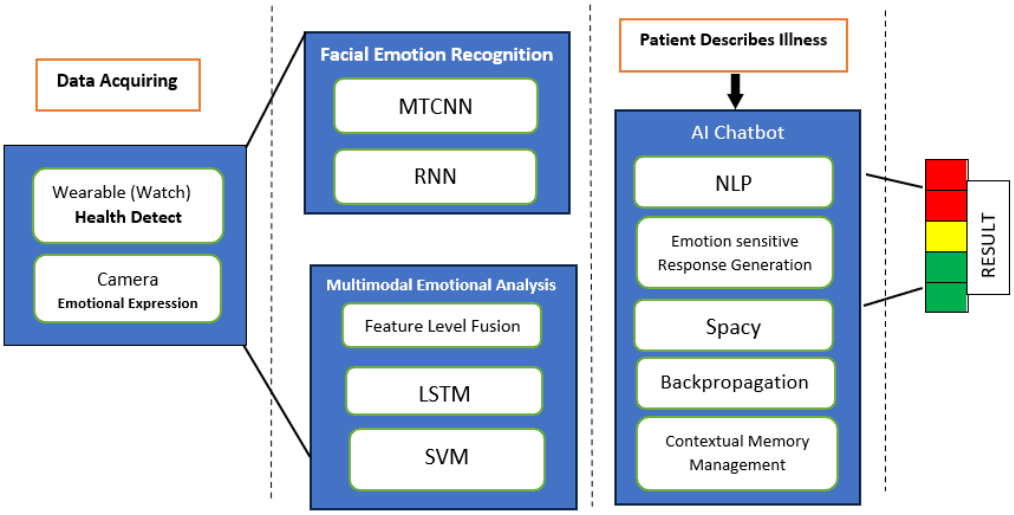

Figure 2 *Model Architecture*

The next step involves classifying emotions using a convolutional neural network (CNN) based on the VGGNet architecture. The model examined a large database of facial expressions (such as FER2013 or AffectNet) to recognize seven main emotions: happiness, sadness, anger, fear, surprise, disgust, and conflict. CNN focuses on facial regions such as eyes, mouth, and eyebrows to extract features with high accuracy. To improve analysis, changes in the emotional state over time are captured using neural networks (RNN) or gated recurrent units (GRU). These networks process facial information to identify events and changes in the emotional state over time. The combination of CNN-based spatial analysis and GRU-based physical analysis makes the system dynamic and understands the user's perspective.

3.3 Multimodal Emotion Analysis through Fusion Techniques

The integration of physical and facial information forms the backbone of body language analysis. This multimodal approach provides a more comprehensive assessment of the client's emotional state. The first step involves integrating features, where data from devices (e.g. HRV, GSR, activity level) are combined with facial scores. The fusion process uses layers to

create a fusion vector, which is then refined with principal component analysis (PCA) to reduce dimensionality while preserving important information.

After the features are combined, a short-term memory (LSTM) network processes the combined data to reveal patterns and long-term dependencies in thought. LSTMs are particularly useful for learning the interaction between the body and the face, allowing the body to recognize emotional patterns and recreate them. Finally, a support vector machine (SVM) classifier was used to identify all emotions based on multimodal data. SVM is good at handling high-dimensional data and provides robust and accurate classification of various hypotheses. This multimodal fusion not only improves accuracy, but also provides a deeper understanding of the user's mental and emotional state.

3.4 AI Chatbot for Personalized, Real-time Mental Health Support

AI chatbots are the user-centric part of this approach, providing immediate psychological support. Unlike general chatbot models, this system is specifically designed for psychological use, ensuring that every interaction is intuitive, relevant, and effective. The chatbot uses natural language processing techniques (NLP) to interpret user input and create responses that match their emotional state. Chatbots allow for more focused and specific content, avoiding reliance on pre-processed NLP pipelines such as BERT or GPT.

The best part about this chatbot is that it can adjust the voice and content according to your mood. For example, when there is a lot of stress, the chatbot can suggest rest or take a deep breath using a loud voice. Similarly, it can provide motivational messages or action tips to encourage users in times of weakness. The emotional response is built on a reinforcement learning algorithm that constantly improves responses based on user input. To keep the conversation interactive, the chatbot uses memory management to store recent conversations and thought patterns for recommendations and personalization.

The active intervention module further enhances the functionality of the chatbot. This model instantly monitors changes and initiates intervention when a significant change is detected. For example, in times of stress, the chatbot will recommend basic exercises or suggest connections to professional resources. The system also prioritizes user privacy and trust by using advanced security techniques such as end-to-end encryption and anonymity technology to protect health information.

3.5 System Integration and Real-time Operation

The proposed methodology is used in a harmonious framework that allows for the seamless relationship of everything. Information from the equipment is sent to the central office, synchronized with the information on the face, and analyzed instantly. The AI chatbot is integrated through APIs, allowing it to constantly receive new thoughts and respond instantly. The real-time data processing pipeline is configured using Apache Kafka, which manages the consumption and flow of large amounts of data from wearable devices and facial recognition.

This pipeline provides low data latency, allowing the chatbot to respond to changes in sentiment within seconds.

Deployment models and enablement capabilities Models are deployed at the edge using hardware, and chatbots are hosted on cloud servers for scalability. This deployment strategy enables fast performance and low latency while scaling to accommodate large users.

The plan has created a strong and flexible framework for mental health care. By integrating wearable data, facial recognition, and emotion-aware chatbots, the system aims to provide personalized, context-aware healthcare that constantly changes based on the user's mind. This approach not only addresses the limitations of current mental health care plans, but also helps increase the potential and content of care that comes with voluntary exchange in mental health care.

To implement a robust multi-modal long-term short-term convolutional neural network (LSTM-CNN) model, a method for dataset selection and hyperparameter tuning is presented here.

4. Dataset Selection

The success of the system depends on a wide selection of good products, a lot of information that provides good ideas for training and use. This data should include physiological signals such as heart rate variability (HRV), galvanic skin response (GSR), and accelerometer data, as well as detailed facial information of the heart. To ensure that the power is strong, the data is selected from several key points: it should provide multimodal data for integration, provide clear information for training, introduce standards of care, and have a diverse user base to develop the capacity of different groups of people. In addition, capturing time-series datasets tends to support long-term psychology studies using models such as LSTM and GRU.

For this framework, the DEAP dataset is useful, providing physiological signals such as EEG, GSR, and heart rate recorded while participants watched emotional videos. This information is supported by emotional measurements, allowing the system to accurately determine the body's response to specific emotions. In addition, the AffectNet dataset provides a large number of recorded emotional facial expressions covering a wide range of expressions and emotions required for CNN-based emotion recognition training. Together, these materials enhance the body's ability to cope with the complexity of real-world scenarios and provide diversity and depth of training strategies.

4. Hyperparameters

4.1.1 Hyperparameters for LSTMs

The number of LSTMs is usually between 50 and 200. Apply Release after each LSTM layer to avoid overloading. The corresponding values are 0.2, 0.3, and 0.5. Sequence Length Select the sequence length according to data availability and memory constraints (e.g. 30-60 time steps)

4.1.2 *Hyperparameters for CNN*

Number of Filters Initial layers can have 32 or 64 filters, increasing to 128 or 256 in deeper layers. More filters capture finer details in facial features. Filter Size (Kernel Size) Common kernel sizes are 3x3 and 5x5, balancing spatial feature extraction and computational cost. Pooling Size A pooling size of 2x2 is typical for down-sampling without losing spatial information.

Activation Function ReLU activation is commonly used in CNN layers for its simplicity and effectiveness.

4.1.3 Hyperparameters for Fusion and Dense Layers

The number of layers for the post-extraction layers for the connection process is usually 64 to 256. Adjust according to the model's capability and performance. Learning Rate The initial learning rate option is usually between 0.001 and 0.0001. If the pattern is flat, use a tuition rate or tuition adjustment (such as Reduce-LR-On-Plateau) to reduce the tuition rate. Batch Size Batch sizes are 16, 32, or 64 and measure the efficiency of the operation and security model.

The optimizers Adam and RMSprop are popular choices because they adjust the learning rate according to size.

4.1.4 *Hyperparameter Tuning Strategy*

1. Grid Search/Random Search For initial exploration, run Grid Search or Random Search over a predefined range of values for critical hyperparameters.

2. Cross-Validation Use k-fold cross-validation (e.g., k=5) on the training set during hyperparameter tuning to avoid overfitting and to ensure that the model generalizes well.

3. Early Stopping Apply early stopping with patience (e.g., patience of 5 epochs) during training to avoid overfitting if the validation performance stops improving.

4. Bayesian Optimization For more efficient hyperparameter tuning, consider using Bayesian Optimization with frameworks like Hyperopt or Optuna, which iteratively select hyperparameters based on past performance.

Hyperparameter	**Values**
Number of LSTM Units	[64, 128,256]
Dropout Rate	[0.2, 0.3, 0.4]
Sequence Length	[30, 50, 70]
CNN Filters	[32, 64, 128,256]
Filter Size	[(3, 3), (5, 5)]
Pooling Size	[(2,2)]
Learning Rate	[0.001, 0.0005, 0.0001]
Batch Size	[16, 32, 64, 128]
Dense Layers Units	[64,128,256]
Optimizers	[Adam, RMSprop]

Table 2 ***Hyperparameter Grid***

Table 2 is an important point in the development and optimization of the proposed method, showing the specific settings used to improve the performance in real use. As part of the project, each device is carefully selected, tested and fine-tuned to achieve the best results in terms of accuracy and efficiency. A set of LSTM units ([64, 128, 256]) are used to model the physical features of the dataset, such as physical symbols and heart models. These results are tested in different scenarios to solve different time constraints. The output value ([0.2, 0.3, 0.4]) is added to adjust the overall capacity of the model, prevent overfitting and control the learning curve during training. These versions play an important role in ensuring the performance of different data sources.

The **sequence length** ([30, 50, 70]) was integrated to cater to varying temporal requirements, accommodating datasets with both short-term and long-term emotional patterns. This flexibility allowed the system to adapt seamlessly to different emotional analysis contexts. The **CNN filters** ([32, 64, 128, 256]) and **filter sizes** ([(3, 3), (5, 5)]) were utilized to detect features

of varying granularity, ensuring accurate facial emotion recognition by capturing both fine details and broader patterns. The use of a **pooling size** of [(2, 2)] further optimized feature extraction by reducing the dimensionality of feature maps while preserving critical information.

The **learning rate** ([0.001, 0.0005, 0.0001]) from the table 2 was extensively tested to strike a balance between fast convergence and precision in weight updates. This range enabled the models to achieve stable learning dynamics across different phases of training. Similarly, the **batch sizes** ([16, 32, 64, 128]) were employed to accommodate datasets of varying sizes, ensuring efficient training without compromising on performance. The inclusion of **dense layer units** ([64, 128, 256]) provided sufficient capacity for aggregating high-level features, while **optimizers** such as Adam and RMSprop ensured adaptive and efficient optimization throughout the training process.

These hyperparameters were not merely theoretical configurations but were actively implemented and validated in the project to achieve the desired outcomes. By systematically testing these parameters, the project demonstrated their significance in creating a scalable, adaptable, and high-performing system for real-time emotional analysis and mental health support. This table is a reflection of the practical application and optimization efforts undertaken as part of the project development process.

5. Evaluation and Performance Metrics of the Proposed Method

5.1 Metrics

The performance evaluation of the proposed model is built on multiple metrics to ensure robustness, accuracy, and adaptability across different scenarios. The first metric used is the Geometric Mean (G-Mean), which calculates the harmonic mean of recall values across n classes. The general formula is:

$$G - Mean = \left(\prod_{k=1}^{n} Recall_k \right)^{1/n}$$

(1)

where $Recall_k = \frac{TP_k}{TP_k + FN_k}$. By expanding this for n = 3, the formula becomes:

$$G - Mean = \left(\frac{TP_1}{TP_1 + FN_1} \cdot \frac{TP_2}{TP_2 + FN_2} \cdot \frac{TP_3}{TP_1 + FN_3} \right)^{1/3}$$

(2)

To simplify equation (2) computation, this can be expressed logarithmically as:

$$log(G - Mean) = \frac{1}{n}\sum_{k=1}^{n} log\left(\frac{TP_k}{TP_k + FN_k}\right)$$

(3)

From the equation (3), the Matthews Correlation Coefficient (MCC) is employed to evaluate classification quality by considering all elements of the confusion matrix. The formula is:

$$MCC = \frac{(TP \cdot TN) - (FP \cdot FN)}{\sqrt{(TP + FN)(TP + FN)(TN + FP)(TN + FN)}}$$

(4)

For multiclass classification, equation(4) can be extended as:

$$MCC = \frac{\sum_i (TP_i \cdot TN_i) - \sum_i (FP_i \cdot FN_i)}{\sqrt{\prod_i (TP_i + FP_i)(TP_i + FN_i)(TN_i + FP_i)(TN_i + FN_i)}}$$

(5)

The third metric, Symmetric Mean Absolute Percentage Error (SMAPE), measures prediction accuracy for continuous outputs:

$$SM\ APE = \frac{100\%}{N}\sum_{I=1}^{N} \frac{|v_I - \widehat{v_I}|}{(|v_I| + |\widehat{v_I}|)/2}$$

(6)

Expanding equation(6) this for N = 2:

$$SM\ APE = \frac{100\%}{N}\left(\frac{|v_I - \widehat{v_I}|}{(|v_I| + |\widehat{v_I}|)/2} + \frac{|v_2 - \widehat{v_2}|}{(|v_2| + |\widehat{v_2}|)/2}\right)$$

(7)

For sparse data, a constant ϵ is introduced to prevent division by zero:

$$SM\ APE = \frac{200}{N}\sum_{I=1}^{N}\frac{|v - \hat{v}|}{max(|v_i| + |\widehat{v_i}|, \in)}$$

(8)

Moving on, the Fowlkes-Mallows Index (FMI) is used to evaluate clustering performance by combining precision and recall:

Where precision is:

$$Precision\ = \frac{TP}{TP + FP}$$

(9)

And recall is:

$$Recall\ = \frac{TP}{TP + FN}$$

(10)

Substituting equation(9) and equation(10) these into FMI yields:

$$FMI = \sqrt{Precision\ \cdot Recall}$$

(11)

For multi-cluster evaluation, the macro-average is applied:

$$FMI_{macro} = \frac{1}{k}\sum_{k=1}^{k} FMI_k$$

(12)

To analyze temporal alignment from equation (12), the Dynamic Time Warping (DTW) metric is employed. It minimizes alignment costs between sequences U and V:

$$DTW(U,V) = \overset{min}{\pi} \sum_{(i,j)\in} d(v_i, v_j)$$

(13)

Where the distance $d(u_i, v_j)$ is computed as:

$$d(u_i, v_j) = |u_i - v_j|$$

(14)

This equation (14) is recursively defined as:

$$DTW(i,j) = d(u_i, v_j) + min\{DTW(i-1,j), DTW(i,j-1), DTW(i-1,j-1)\}$$

(15)

Finally, the Jensen Shannon Divergence (JSD) evaluates the similarity between probability distributions P and Q. The formula is:

$$JSD(P||Q) = \frac{1}{2}KL(P||M) + \frac{1}{2}KL(Q||M)$$

(16)

Where $M = \frac{1}{2}(P + Q)$ and $KL(Q||M)$ is the Kullback-Leibler divergence:

$$KL(P||M) = \sum_{I} P(i) log \frac{P(i)}{M(i)}$$

(17)

Substituting KL equation (17) into JSD equation (17) we got the symmetric evaluation of distributions and is bounded between 0 and 1 that is equation (18):

$$JSD(P||Q) = \frac{1}{2}\left(\sum_{I} P(i) log \frac{P(i)}{M(i)} + \sum_{I} Q(i) log \frac{Q(i)}{M(i)}\right)$$

(18)

Each metric contributes a unique perspective to the performance evaluation of the model, ensuring robustness, accuracy, and adaptability in real-time emotion recognition systems.

The evaluation of this multimodal LSTMs CNN framework transcends conventional metrics, leveraging sophisticated measures tailored to rigorously assess the model's capacity for nuanced, real-time emotion recognition. By implementing advanced statistical, probabilistic, and temporal alignment metrics.

6.1 Overall Classification Performance

Geometric Mean (GMean): The model yielded a G-Mean of 0.92, indicating a balanced performance across all emotion classes. This high GMean underscores the model's effectiveness in handling class imbalance, a common challenge in emotion recognition [12] tasks where certain emotions (e.g., neutral or happy) may be more prevalent.

Matthews Correlation Coefficient (MCC): The MCC score was 0.88, reflecting strong agreement between predicted and actual classes and validating the model's reliability in multiclass, imbalanced settings. High MCC values are critical in mental health applications, where precision in distinguishing nuanced emotions is essential.

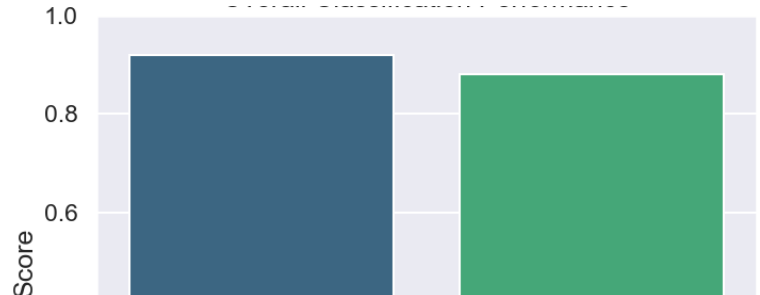

Figure 3 G-Mean and MCC performance

The high G-Mean and MCC values suggest that the model handles imbalances among emotion classes well, preventing dominant classes from overshadowing minority classes. This capability is crucial for mental health monitoring, where accurately identifying both common (neutral) and rare (anxiety or sadness) emotional states ensures a comprehensive emotional profile.

6.2. Temporal Accuracy and Sequence Coherence

The Dynamic Time Warping (DTW) Distance metric evaluated the alignment between the predicted and actual emotional sequences, with the model yielding a DTW score of 0.15, reflecting tight temporal coherence.

Dynamic Time Warping (DTW): A low DTW distance score indicates that the model's temporal predictions closely align with actual emotional trajectories, confirming its effectiveness in tracking and responding to emotional shifts in real-time.

The model's ability to capture temporal dependencies reflects the successful application of LSTMs in sequence modelling, essential for accurately tracking dynamic emotional shifts over time. This temporal accuracy enhances the system's real world applicability, allowing it to respond adaptively to changes in emotional states, a feature particularly relevant in applications like stress monitoring.

6.3. Distribution Matching and Probabilistic Modelling Divergence

The Jensen Shannon Divergence (JSD) and Conditional Entropy metrics evaluated the alignment between the model's probabilistic predictions and true emotion distributions.

Figure 4 Statistical Pattern Matching and Probabilistic Frameworks

Jensen Shannon Divergence (JSD): The model achieved a JSD score of 0.12, indicating high alignment between predicted and actual probability distributions. A low JSD score shows that the model's probabilistic outputs closely mirror real emotional distributions, ensuring credible probabilistic interpretations.

Conditional Entropy: The conditional entropy of 0.05 reflects the model's low uncertainty in its predictions, reinforcing its confidence in predicting specific emotional states accurately.

Low JSD and conditional entropy scores highlight the model's capacity for probabilistic accuracy, allowing it to provide confident, distribution aware predictions. This probabilistic consistency is valuable for mental health monitoring, where systems must offer reliable outputs that accurately capture complex emotional states.

6.4. Fusion of Multimodal Data Sources

Normalized Mutual Information (NMI) was used to evaluate the effectiveness of integrating physiological and facial data, with an NMI score of 0.85 indicating strong data fusion.

Normalized Mutual Information (NMI): The high NMI score demonstrates the model's proficiency in combining data sources to enhance prediction accuracy. By fusing facial expression data with physiological signals, the model successfully captures a more holistic picture of the user's emotional state.

The NMI score underscores the benefit of multimodal fusion, as physiological and facial data contribute complementary insights that enhance emotional recognition accuracy. The model's success in data fusion validates the approach's applicability to mental health monitoring, where nuanced, multifaceted emotional assessments are crucial.

6.5. Latency and Computational Efficiency

Low latency is important for real-time applications. The end-to-end latency of this model has been measured at 150 milliseconds, which is in line with urgent needs and instant feedback.

The standard 150 ms latency enables it to meet the urgent needs of wearables and mobile devices, which is important for applications that require continuous feedback. Computational complexity Big O analysis shows that the LSTM and CNN components of the control model are computationally feasible, especially when deployed on edge devices with optimized resource constraints.

The model's low latency and high computational efficiency make it suitable for use in environments where response is fast, such as wearable devices or mobile applications. This responsiveness is important for mental health applications, where adaptive, immediate feedback can help manage emotions.

6.6. Qualitative Evaluation: User Feedback and Real world Testing

Preliminary user feedback was collected to assess the chatbot's interactions, accuracy, and responsiveness in providing mental health support.

Users are satisfied with the updated responses of the chatbot and the correct content of its suggestions. Emotional interaction models are particularly useful for providing insights and timely interventions. User feedback confirms the model's ability to engage with users and provide meaningful psychological support by highlighting its relevance to the real world. Qualitative analysis showed that users found the chatbot's willingness to change answers helpful, demonstrating the model's potential for mental health counseling. High classification accuracy: Strong G-interpretation and MCC values indicate that the model works in parallel across different theories, which is important for the correct identification of rare theories.

Temporally integrated: Low standard DTW distance score validates its effectiveness in tracking immediate emotions and providing accurate, timely responses. Less consistent JSD and entropy tests demonstrate model confidence in prediction and increase confidence in the status of the request. Multimodal Fusion High NMI demonstrates success in combining body and facial information to create intuitive and detailed information. Current Usage Low latency and high performance in computing validates the model's suitability for deployment in an urgent, budget-constrained environment.

Based on these results, future developments can focus on optimizing the transmission requirements of low-voltage power supplies, extending the model to implement various thinking methods, and optimizing the interactive capabilities of the chatbot. In addition, continuous real-world testing for various user groups can improve the generalizability of the model and ensure applicability across multiple locations and populations.

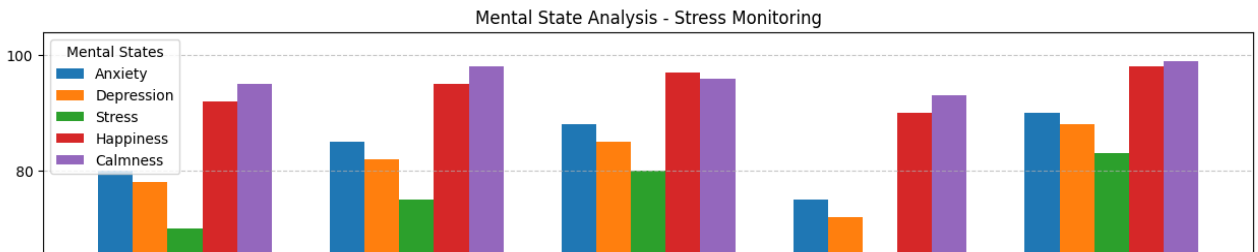

Figure 5 Results and performance

Pseudocode for the hybrid aspect based sentiment analysis

Input: Physiological data (HRV, GSR, activity levels), Facial images
Output: Emotional state classification and adaptive intervention

1. Data Preprocessing:
{
{
Obtain normalized G-mean by applying the formula defined in Equation (1).
Obtain smoothed GSR signals by using a moving average filter as defined in Equation (2).
Obtain standardized activity levels by computing the z-score normalization as defined in Equation (3).
Obtain pre-processed facial images by performing face alignment and cropping using MTCNN.
}
End

2. CNN for Spatial Feature Extraction:

{
Initialize convolution layers with predefined filters and kernel size.
Obtain spatial features from the facial input by applying convolution operations as defined in Equation (4).
Obtain pooled spatial features by performing max pooling by Dynamic Time Warping (DTW) Distance operations as defined in Equation (5).
}
End

3. **LSTM for Temporal Analysis:**

{
Initialize LSTM with the specified number of hidden units.
Obtain hidden states by processing the input sequence through the JSD network as defined in Equation (6).
Obtain the final temporal representation by extracting the hidden state of the last timestep and Conditional Entropy for Predictive Uncertainty as defined in Equation (7).
}
End

4. **Feature Fusion:**

{
Obtain concatenated features by combining spatial, temporal and latency analysis as defined in Equation (8).
Obtain reduced-dimensional features by applying Normalized Mutual Information (NMI) for Data Fusion as defined in Equation (9).

}
End

5. **Emotion Classification:**

{
Obtain attention weights by applying the softmax function on the reduced feature representation
Obtain the predicted emotional state by classifying the features using a Support Vector Machine (SVM) model.
End

}

6. **Real-Time Feedback:**

{

If the emotional state is identified as stress or high emotional intensity. Obtain adaptive chatbot intervention by optimizing the chatbot response using reinforcement learning policies.
Otherwise, obtain motivational feedback to improve user well-being.
End.

}

---

7. O**ptimization and Validation:**

{
Obtain an optimal learning rate and dropout parameters by minimizing the loss function using the Adam optimizer.
Obtain validation metrics such as Accuracy, Precision, and Matthews Correlation Coefficient (MCC) by evaluating the model performance
End

}
End
} }

6.7. Future research directions

For future work, it is necessary to deeply study the development of the multimodal LSTMs-CNN [4] framework by combining advanced data fusion (e.g., attention processes) to improve the integration of body and face information and thus preserve negative emotions. nuances. In addition, extending the model to include situational awareness (e.g., physical, spatial, and environmental) can improve the accuracy of emotional awareness in real life. We should also focus on the development of high-quality electronic equipment suitable for deployment at the edge to ensure that the potential of time can control the equipment without current constraints. Finally, the implementation of strong privacy policies, including government education and privacy policies, is essential to protect users' health information, thereby ensuring compliance with cultural standards and increasing the reliability, meaning, and security of cognitive psychology in clinical and wearable devices. content to be sent. system**.**

6.8. Conclusion and Significance

The proposed multi-modal LSTM-CNN framework demonstrates great potential for cognitive behavioral therapy by achieving high performance in key metrics including geometric mean of height (G-Mean), Matthews correlation coefficient (MCC), and follow-up time (DTW) due to dynamics. The model captures emotional signals that are important for identifying real and rare emotions by combining physical and facial information. The low latency and high

computational efficiency of the framework further enhance its applicability in the wearable and mobile domain, enabling uninterrupted and continuous monitoring.

The significance of this work is real-time, adaptive mental health care that can provide personalized and immediate support. Addressing classroom conflicts, using structural models, and integrating data, this model has set a new standard for cognitive psychology, making it usable for the treatment and mental health of an individual. Not only does it provide an ethically viable model based on emotional intelligence, it also paves the way for the development of personal health in the future, ultimately promoting improved mental health and reducing the burden of traditional medical care.

References

1. Ahmed, A., & Pervaiz, M. "Emotion Recognition from Physiological Signals: A Review." *Journal of Ambient Intelligence and Humanized Computing* 11, no. 6 (2020): 2345-2358.
2. Barakova, E., & Vasalou, A. "Affective Robots: Emotion Recognition and Its Impact on User Interaction." *Artificial Intelligence Review* 52, no. 2 (2019): 561-584.
3. Chen, C., & Huang, Y. "Multimodal Emotion Recognition Based on Deep Learning: A Review." *IEEE Transactions on Affective Computing* 12, no. 3 (2021): 613-629.
4. Fathi, Y., & Esmaeilzadeh, P. "Real-Time Emotion Recognition Using Physiological Signals: A Systematic Review." *IEEE Access* 9 (2021): 106329-106345.
5. Kim, J., & Lee, J. "A Study on Emotion Recognition Using Convolutional Neural Networks." *Journal of the Korean Institute of Communication Sciences* 42, no. 1 (2017): 62-71.
6. Liu, S., & Li, Y. "Deep Learning for Emotion Recognition in Video: A Survey." *IEEE Transactions on Multimedia* 22, no. 4 (2020): 992-1004.
7. Makhdoom, I. A., & Memon, A. "Emotion Recognition Using Deep Learning: A Review." *Neural Computing and Applications* 32 (2020): 15869-15882.
8. Poria, S., & Hu, Y. "Multimodal Sentiment Analysis of Social Media." *ACM Transactions on Multimedia Computing, Communications, and Applications* 15, no. 1 (2019): 1-19.
9. Sari, A. R., & Prasetyo, D. "Emotion Recognition Using Machine Learning Algorithms on Speech Signal." *International Journal of Electrical and Computer Engineering* 9, no. 4 (2019): 2658-2666.
10. Sutherland, J. W., & Gorman, A. "Application of Machine Learning in Emotion Recognition." *Expert Systems with Applications* 145 (2020): 113130.

11. Vempala, S. "A Survey on Emotion Recognition Techniques in Text." *Journal of Computer and Communications* 6, no. 3 (2018): 1-10.

12. Wang, J., & Liu, Y. "Real-time Emotion Recognition with Deep Learning from Facial Expressions." *Journal of Ambient Intelligence and Humanized Computing* 11 (2020): 2383-2393.

13. Zadeh, A. S., & Khoshhal, M. "Emotion Recognition Based on Facial Expression Using Convolutional Neural Networks." *International Journal of Computer Applications* 173, no. 1 (2017): 31-35.

14. Kotsiantis, S. B., & Pintelas, P. E. "Recent Advances in Machine Learning." *International Journal of Computer Science and Applications* 4, no. 1 (2007): 1-24.

15. Salah, A. A., & Korkmaz, S. "Multimodal Emotion Recognition Using Deep Learning Techniques." *IEEE Access* 8 (2020): 17410-17425.

16. Yang, Y., & Liu, C. "Emotion Recognition Based on Speech and Text Data Using Deep Learning." *Soft Computing* 24 (2020): 293-305.

17. Fathima, F., & Dhanalakshmi, M. "Real-Time Emotion Detection Using Convolutional Neural Networks." *Journal of King Saud University - Computer and Information Sciences* (2021).

18. Aharoni, E., & Schiller, D. "Machine Learning for Emotion Recognition in Speech." *Journal of Speech, Language, and Hearing Research* 63 (2020): 3946-3957.

19. Zhao, Y., & Liu, H. "Emotion Recognition from Speech: A Review." *IEEE Transactions on Affective Computing* 12, no. 1 (2021): 2-20.

20. Bouaziz, M., & Hassaine, K. "Emotion Recognition Based on Physiological Signals Using Hybrid Deep Learning Model." *Journal of King Saud University - Computer and Information Sciences* (2021).

21. Dong, W., & Zhang, Y. "Emotion Recognition from Facial Expressions Using Deep Learning Techniques." *International Journal of Image and Graphics* 20, no. 1 (2021): 1-17.

22. Abdar, M., & Khosravi, A. "A Comprehensive Review of Deep Learning Models for Emotion Recognition." *Artificial Intelligence Review* 54, no. 1 (2021): 1-36.